THOUGHTS ABOUT A SOLUTION TO THE MIND-BODY PROBLEM

Arnold Zuboff

This challenging paper presents an ingenious argument a for functionalist theory of mind.

The Phenomenal Key

Please try seeing figure A below (called by perceptual psychologists a 'Necker cube') first as suggested by figure B and then as suggested by figure C. First a slightly irregular transparent cube seems pictured as though we are looking down on it, with the lower square in figure A seen as the front of the cube. Then such a cube seems pictured as though we are looking up at it, with the lower square in figure A seen this time as the back of the cube. Try making the look of the picture shift back and forth between these two.

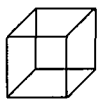


Figure A

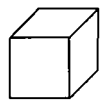


Figure B

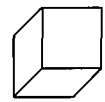


Figure C

Now, I believe that there are two levels of naivety we may naturally find ourselves at in an attempt to understand what is happening in perception. The Necker cube helps to expose naivety at both levels.

Here is the first: For the most part, we naturally think that in perception we are merely sponges soaking up the various qualities of the things we perceive, qualities that would be just as we perceive them to be quite apart from their being perceived. I open my eyes, and before me are things merely showing to me the qualities they wear on their faces. And that this is so is, quite simply, how it looks to me. The problem for that view in this special case is that, though it looks that way here too, we can know that at least in this case that look is deceptive. For, though it looks as though there is a change occurring over there, on the paper – a shifting in the look of the picture – in this case the perceiver knows that the only

https://doi.org/10.1017/S1477175600003109 Published online by Cambridge University Press

real change is rather in the perceiver. This is a change that the perceiver himself is bringing about in how the object is being seen. The shifting look of the picture is thus revealed to be a product of interaction between the object and the subject's perceptual activity, with the latter being crucial for what the look of the object will be.

But what is it that is crucial within the perceptual activity itself? Let me try to express this question more precisely: We have been discussing the look of orientation in an object of perception, a picture. And we have come to recognise that looking like this is a phenomenal property, belonging to the picture by virtue of how it appears to us. What we have also come to recognise, then, is that there must be some corresponding property in the perceiver, which could also be called phenomenal, that property by virtue of which the object appears to him as having *its* phenomenal property. Calling both of these paired properties phenomenal must not suggest, of course, that the phenomenal property in the subject appears for that subject as does the corresponding phenomenal property in the object. The subject is not directly aware of what it is in him that constitutes his awareness of the object. What is shown him in the experience is the object, not the state within himself which is the experiencing of that object. Well, the question we are interested in is this: What, more specifically, is this phenomenal property in the perceiver? The answer to this question, I believe, is the key to an understanding of the mental. And it is regarding this that I think the second level of naivety needs exposing.

Three areas

It would be useful to distinguish three areas, as we might call them, containing the things and activities that are involved in or associated with perceptual experience: The first, the object area, which is external to the perceiver except in cases like bodily sensation, contains the experienced object and perhaps also a medium through which the object acts on the sense organs, as does light in vision. We have dismissed this area

https://doi.org/10.1017/S1477175600003109 Published online by Cambridge University Press

as the focus of our interest. The second area, the processing area, contains the perceptual processing that is generated within the perceiver's brain through the stimulation of the sense organs. The third area, the response area, contains the pattern of readiness to respond to the object on account of how it is perceived, and its embodiment is in parts of the brain like the speech centre and the centre for voluntary movement as well as parts of the brain in which thinking about the object and the experience is localized.

Let's examine more closely what is happening in the second area of perception, the processing area. In vision the pattern of light that falls on the retinas generates neural impulses that with amazing intricacy and thoroughness are, one might say, filtered – first on the way to and then within the visual cortex at the back of the brain – variously filtered according to countless more or less general features of the pattern on the retina. For example, there will be a neuron made to fire by all, and by only, right-angled patterns no matter where they occur on the retina. Obviously our perception of the right angle as a general property must in some way depend on such specialized neural firing.

Then comes a transition from this second area, this processing area to the third area, the response area, as further transmissions of neural impulses fan out from the visual cortex into other parts of the brain; and in these parts will be fixed the ways we are ready to behave and speak and think based on how the Necker cube looks.

Intrinsicism

The completely natural temptation that I am claiming is naive is to think that the look of an object is already established in the processing area, in the visual cortex. The temptation is to think that some bit of processing in the visual cortex possesses within itself the phenomenal property we are seeking. I am going to argue that what fixes a look is rather the pattern of readiness to respond to the look. (Let me point out an essential subtlety here: My view is that what might best be thought of

https://doi.org/10.1017/S1477175600003109 Published online by Cambridge University Press

as in some sense possessing the phenomenal property may indeed be a relevant bit of processing in the visual cortex, but that such processing could have no phenomenal property in itself. It would possess its phenomenal property only on account of its causing the appropriate responsiveness to the corresponding phenomenal property, the look, of the object.)

But doesn't my view seem hopeless from the start? After all, it just seems obvious that a look is established first, that only *after* a look has been established can we become appropriately responsive to that look. The responsiveness must be caused by the experience of the look to which it is responsive, and therefore the experience must possess the character through which it excites this responsiveness independent of the responsiveness itself. Now, in physiological terms what causes the responsiveness to a look is something happening in the visual processing, in the visual cortex. So the natural view is that something in the intrinsic character of that bit of the processing is what crucially determines the way that something looks. I shall call this view 'intrinsicism'.

There are two competing versions of this natural view. One is the physicalist version. Those many physicalists who are also intrinsicists will identify, say, the experience of seeing a red triangle with that bit of the processing in the visual cortex that 'plays the causal role' of the experience. That is, they will identify the experience with that bit of the processing which is caused by the causes of the experience of seeing the red triangle and in turn is itself the cause of whatever is an effect of the experience of seeing the red triangle. Now, I too shall be making exactly that identification although I am arguing against intrinsicism. Here is the difference between my view, functionalism, and a physicalist's intrinsicism: The functionalist thinks that playing its causal role is all that gives that bit of neural activity its mental character, including its phenomenal nature, whereas the intrinsicist thinks that properties intrinsic to that bit of neural activity, which could be chemical or even biological, must be essential to the subject's experiencing of the phenomenal qualities of the object. So, for the physical intrinsicist, the phenomenal property in the subject – that which

https://doi.org/10.1017/S1477175600003109 Published online by Cambridge University Press

is essential to the subject's experiencing of the phenomenal property in the object – is some such intrinsic physical property in the relevant bit of processing.

But some intrinsicist philosophers understandably balk at identifying a phenomenal property with any of these purely physical intrinsic properties of neural activity. It seems clear to them (and to me) that a phenomenal property (that whose involvement in some state of the perceiver would not cause but simply *be* the perceiver's being conscious in the way in question) is necessarily unidentifiable with anything like chemical or biological properties. Consider it: The merely physical event of an electro-chemical impulse surging through a neuron that is hooked up to fire when there are right angles on the retinas is supposed to have what it takes, *within itself*, on account of its chemical or biological nature, to *be* the *experiencing* of something as a right angle? But there is nothing within that neuron that is particularly redolent of right angles – or of experience. It has within itself, as philosophers say, no 'intentionality', no intimate relating of itself to the object of the experience, to a right angle. (Let me quickly mention that this neuron's *extrinsic* property of causing responsiveness elsewhere in the brain that is appropriate specifically to the seeing of a right angle, this might much more easily be supposed to give the firing of this neuron intentionality.)

Those intrinsicists who thus balk at identifying a phenomenal property with any purely physical intrinsic properties of neural activity will therefore reject physicalism, the view that there are only physical entities and properties, in that they assert that there must be two irreducibly distinct sorts of properties involved in perception, the physical and the mental. These days such philosophers are unlikely to be substance dualists and interactionists like Descartes. They will probably accept the scientific evidence that Descartes did not have that there are no causal gaps in the physical workings of the brain of a sort that would have to be discoverable if the mind was an entity that was distinct from the brain and interacting with it. So these philosophers are property rather than substance dualists. They think that the brain is the entity that possesses the irreducible

https://doi.org/10.1017/S1477175600003109 Published online by Cambridge University Press

mental properties along with its physical properties.

And they are epiphenomenalists rather than interactionists. For the epiphenomenalist there is a one-way causation of the mental by certain physical properties. The mental properties of experience are generated by the relevant physical properties of the brain; but these mental properties don't themselves have any effects on the physical brain (because otherwise causal gaps in the physical activities of the brain would have been discovered). The phenomenal events are epiphenomena of the brain events. And, as intrinsicists, these philosophers believe that the phenomenal characteristics of experience are *generated by* (and therefore not identical with) the very sorts of intrinsic physical properties in the perceptual processing with which the physicalists had unconvincingly *identified* them.

The views

Here is a brief exposition of the views I have mentioned: The light from a book on a shelf reaches my retinas. The neural processing is done behind the eyes, on the way to and then within the visual cortex. The interactionist dualist thinks that the physical pattern thus produced in the brain somehow, mysteriously, causes the visual experience of a book to occur in a mind, a self, which is a non-physical entity, without parts, without extension, and, of course, distinct from the brain, with which it is nevertheless interacting. Within the mind the experience of the book, let us say, leads to a decision to reach for the book. The will within the mind consequently acts, again in mysterious fashion, on a part of the brain that causes impulses to be sent to the muscles involved in taking hold of the book; and so the action is performed. The causal gap in the working of the brain that interactionist dualism must predict, and that science has shown does not exist, would come, in this example of vision, between the visual processing and the beginning of the neural signals to the muscles. For the non-material mind, not directly detectable by physical means, should be undetectably carrying events forward from the experience of the book to the decision to reach for it and then to the willing that finally causes the brain to excite the muscles.

https://doi.org/10.1017/S1477175600003109 Published online by Cambridge University Press

The physical intrinsicist, the epiphenomenalist and the functionalist all agree that there are no causal gaps of the sort predicted by the interactonist. The light from the book falls on the retinas. The visual processing occurs. This in turn, through neural transmission, brings about appropriate effects within the parts of the brain that are responsible for making decisions, for voluntary movement and so on. The centre for voluntary movement excites the muscles that would be involved in getting hold of the book; and so the action is performed. The physical intrinsicist will identify the phenomenal character of the visual experience of the book with intrinsic physical characteristics of the processing in the visual cortex. The epiphenomenal intrinsicist, will say that these same intrinsic physical characteristics of the processing, while they cannot be identified with the phenomenal properties, generate as epiphenomena the non-physical phenomenal properties involved in seeing the book. The functionalist, however, believes that the phenomenal character of seeing the book, though we may in some sense say it belongs to whatever happens to be playing the causal role of seeing – and therein *is* seeing – which is probably some bit of processing in the visual cortex, belongs to that seeing purely by virtue of its causing the responsiveness in the other parts of the brain that is appropriate to having that visual experience.

Difficulties of intrinsicism

Has the epiphenomenalist really escaped the absurdity of the physicalist identity claim? The epiphenomenalist rightly insisted that the electro-chemical impulse surging through the neuron that is hooked up to fire when there are right angles on the retinas could never be thought of as having what it takes *within itself* to *be* an experiencing of a right angle. Well, why should we think that the electro-chemical impulse somehow has what it takes within itself to *cause* an epiphenomenal experiencing of a right angle? As I pointed out before, there is nothing within that neuron that is particularly redolent of right angles or of experience. Against both its identification with and its generation of the right-angle experience let me add

https://doi.org/10.1017/S1477175600003109 Published online by Cambridge University Press

that internally this bit of neural activity is likely to be virtually indistinguishable from a good deal of the rest of the brain's neural activity. A neuron hooked up to retinal stimulation from red things won't have to differ internally from this right-angle neuron in a way that would be in the least plausible as responsible for the phenomenal difference in the experiences of red and right angles. (Where they will relevantly differ is in the patterns of effects they bring about in the areas of the brain controlling the sort of behaviour, speech and thought that would arise from experiencing red or a right angle.)

And epiphenomenalism loses an important feature of the mental. Surely the phenomenal character of my experience of a red triangle does have causal powers. For example, it is that phenomenal character itself that gets me as a philosopher to talk about that phenomenal character. A powerless epiphenomenon could have no effect on my speech centre and so could never influence what I say. It might seem that the physical intrinsicist has an advantage here. Although it might turn out, as the epiphenomenalist insists, that it is contradictory and therefore impossible to identify the mental with an intrinsic physical characteristic, at least, one might think, such physical characteristics are not epiphenomenal and could therefore have effects like that I've mentioned on speech. But there really is no such advantage because there is no way that the intrinsic character of a bit of visual processing could reach over to my speech centre, somehow apart from its causal role, to affect what I say. Bits of visual processing can affect speech and other responses to experience only by virtue of playing their causal roles. Local intrinsic determinations are as powerless as epiphenomena when it comes to making any specific impression on my responses to experience.

And this problem for the intrinsicist – the causal isolation of intrinsic physical properties – is reflected on the other side of his identification – in a causal isolation of the phenomenal. He believes that the experience of the look of the red triangle is already fully present within the intrinsic state of the visual cortex. What he must think is happening, then, if this already fully realized experience is what is getting me to talk about

https://doi.org/10.1017/S1477175600003109 Published online by Cambridge University Press

having it, is that its phenomenal character is conveyed into my speech by way of the transmission of neural impulses that occurs between the visual cortex and the speech centre. But how can the phenomenal travel through neural impulses?

Let's return for a moment to the example of the Necker cube. For the functionalist the change in the Necker cube's look occurs essentially in a shifting between two patterns of readiness to deal with the orientation of the pictured cube. Seeing the cube one way has you ready, if called upon, to point to a particular square as the front of the cube, imagine something sitting on top of it as having a certain tilt, anticipate the feel of the cube, speak about it, write about it – all of these in a fashion consistent with seeing just one of those orientations; and then somehow you bring about a shift to the pattern of your readiness that is appropriate instead to the other look. (This might be achieved by way of somehow bringing about a change in a relevant bit of processing in the visual cortex such that it would serve as coordinating cause of the required change in the pattern of responsiveness. But it would still be only the change in responsiveness, and not any change in the processing, that was essential to the change in the look.)

The intrinsicist, of course, agrees that there will be a change in the responsiveness that comes with the change in the look. But he understands the responsiveness as related contingently to the experience of the look. The shift in responsiveness, for him, would be an effect of the independently established change in the look. The functionalist, however, says that if there were not this shift in responsiveness it would be logically impossible for the look to have changed. It is a contradiction for you to be ready in all ways to treat the look as if it were of one of these orientations while in fact you are actually seeing it as of the other orientation. And if there is in you a particular one of those patterns of readiness to respond, then you must, necessarily, be experiencing the look to which that responsiveness is appropriate.

Then what of the argument for intrinsicism that I earlier described as a basis for the temptation to embrace it? It seemed obvious that the character of the look of the Necker

https://doi.org/10.1017/S1477175600003109 Published online by Cambridge University Press

cube must have been established before we could develop our readiness to respond to that look. It seemed, therefore, that the experience must possess the character through which it could excite that responsiveness independent of the responsiveness itself.

But recall that from the look of it the shift of orientation in the Necker cube was something happening out there on the paper and, moreover, that event on the paper was causing our experience of it. The peculiar involvement of the will in the case of the Necker cube, however, allowed us to recognise the logical dependence of the phenomenal property of the picture on the very perception that it was supposed to be causing. That property of the object only existed through the existence of the corresponding property in the perceiver. Could it not have been argued that the look in the object just had to be settled independent of the perception of it in order for it to be giving the perception its perceptivity regarding that look? But that would have been a very natural mistake. The objective state of an object usually is a cause of a perception of it (though the Necker cube's orientation is an illuminating exception); but this is by way of effects on the sense organs and not through the possession already of a phenomenal character. Rather, the object depends logically for its phenomenal character on the character of the perception. I am claiming that the same point must now be applied within perception to the relationship between the perceptual processing and the responsiveness that interprets it, that makes the perceiver aware of it, that gives it its phenomenal character. (So the naivety about the object and the naivety about the processing are really just two levels of one and the same naivety.) The look of the Necker cube's orientation, which depends on nothing but our responsiveness to it, is a look that powerfully *seems* fixed independent of that responsiveness (as well as seeming to be so fixed in the object and not the perceiver). But only the naive will let that powerful seeming decide their understanding of perception.

https://doi.org/10.1017/S1477175600003109 Published online by Cambridge University Press

The replacement argument

I shall end this discussion with a brief presentation of an argument for functionalism that I think is decisive, but I would like to preface this with a more general statement than I have so far made of the view for which I am arguing.

The version of the functionalist view of the mind that I support has it that a mental item, such as a belief or desire or sensation, possesses all of its mental character purely by virtue of its causal role in the overall mental system. For example, the experience of red, including in it essentially the peculiar way red feels, would be that experience of red purely on account of its being caused by what causes that experience along with its in turn causing in the rest of the mind and in behaviour whatever that experience would cause. Thus if some purely physical event or state in a brain was playing that causal role of the experience of red, with other events or states in the brain playing the causal roles of the other mental items, then that brain event or state would therein simply be that experience of red. For therein that brain item would have all of the feeling and any other mental characteristics belonging to the mental item. But none of these mental features would be due to any intrinsic features of that brain event or state but rather only to its playing the relevant causal role. Thus, according to functionalism, if a gadget of wires and chips were to replace neurons involved in the playing of that role, then if the gadget were indeed maintaining the same relevant causal relation to the rest of the brain the experience of red would have been fully preserved.

And here is the argument that I earlier said I think is decisive, which I call the 'replacement argument'. I shall present it in the form of a story. (For much more discussion of this argument, see my paper 'What Is a Mind?', in *Midwest Studies in Philosophy*, vol. XIX.)

Imagine that I was carefully frozen just after dying and now find that I am awakened in a time when medical science can restore me to life. The doctors on reviving me tell me that they will replace with a gadget my left visual cortex (which processes the right side of my field of vision). The gadget

https://doi.org/10.1017/S1477175600003109 Published online by Cambridge University Press

replacement will be a precise functional equivalent, perfectly producing the same pattern of impulses to the rest of the brain as the left visual cortex would have, but, of course, with a radical change to the intrinsic properties. On account of my natural intrinsicist prejudice, I complain that it seems to me that this must mess up that side of my vision – perhaps leaving me blind on that side or making things there look metallic or some such. 'Anyway,' I say, 'why do this? My vision is perfect.' But the doctors explain that I've misunderstood. That part of my brain had been badly damaged in my fatal accident, and they were telling me that they had already done the replacement, before I was awakened. So I thank the doctors for my life and my perfectly prserved vision and live happily ever after.

Of course this was a functionalist telling of the story. In it all that mattered to the phenomenal character of experience was the functional. Yet, as I shall point out, there can be no other coherent way of telling it. For the two intrinsicist attempts at a version – 'A Change Expressed' and 'A Change Unexpressed' – are impossible tales.

In 'A Change Expressed', I have been complaining ever since I was awakened with the replacement that the right side of my vision was screwy. But this is no good because our story must be one in which the gadget actually is a perfect functional equivalent. After all, only within such a story could intrinsicism score if the experience was nevertheless different. But therefore anything I say, do or think could not be allowed to be any different from what it would have been with the left visual cortex. So no change in experience could be expressed.

It would seem, then, that the only proper intrinsicist version of the story would have to be 'A Change Unexpressed'. In this I am supposed to be speaking, thinking and behaving as though my vision on the right side were fine even though it is actually all screwy on account of the replacement. But if my vision were so messed up, it would be impossible for me to carry on just as I would if I saw everything as usual on both sides of my field of vision. I'd be bumping into things on the

https://doi.org/10.1017/S1477175600003109 Published online by Cambridge University Press

right and pleading for help from the doctors. No. The only coherent way for all the speech, behaviour and thought of normal vision to have been preserved with the replacement would be for all the experience on which these were based to have been preserved as well. So functioning preserves the phenomenal and functionalism is right.

Arnold Zuboff , Department of Philosophy, University College London.

https://doi.org/10.1017/S1477175600003109 Published online by Cambridge University Press